Applications of object-based image analysis results for the farmland surrounding Kakamega Forest in western Kenya

G. Schaab *,T. Lübker, T. Lung , S.M. Nthuni

Faculty of Geomatics, Karlsruhe University of Applied Sciences, Moltkestraße 30, D-76133 Karlsruhe – gertrud.schaab@hs-karlsruhe.de

Abstract – For 473 km² of farmland surrounding Kakamega Forest in western Kenya, QuickBird satellite imagery has been analyzed by an object-based image analysis approach. Preprocessing involved atmospheric/orographic correction as well as mosaicing and was followed by ground truthing and visual interpretation. Segmentation was optimized using a newly introduced 'area fitness rate' as a discrepancy method and an 'objective function' as a goodness method. The final rule set for classification consisted of 831 individual processes and has resulted in the distinction of 15 land use/cover classes. This wealth of information has provided a thorough basis for a) the analysis of land use and landscape structures leading to ten distinct farmland types, and b) the redistribution of census population data via the development of a GIS-based population surface model. The typology and the QuickBird derived houses or the redistributed population have been used to simulate c) alternative futures of rural livelihood considering price development and crop yields, and d) rainwater harvesting potential.

Keywords: agricultural use, landscape structures, farmland typology, gridded population, socio-economic scenarios, rainwater harvesting potential

1. INTRODUCTION

Within the BIOTA East Africa project, funded by the German Federal Ministry of Education and Research (BMBF) from 2001 to 2010, three forests in Kenya and Uganda were investigated. This provided the means to compare influences of different levels of disturbances, fragmentation, and human use on the biodiversity of East African rainforests (Schaab *et al.*, 2010). Kakamega Forest became the focus area of research and served as a case model when looking into recommendations for a sustainable biodiversity management (Schaab *et al.*, 2009). Here, the conflict between the needs of the rural population and the conservation of biodiversity was the main target of BIOTA research (Schaab *et al.*, 2010).
Kakamega Forest is placed in one of the world's most densely populated rural areas, with an average population density of 643 people/km² (for 1999, projected to be 892 people/km² in 2009) in the 2 km zone adjacent to the reserve's boundary (Schaab *et al.*, 2009). Being claimed to represent the easternmost relic of the Guineo-Congolian rainforest (Kokwaro, 1988), the 238 km² large reserve inhabits species nowhere else to be found in Kenya together with a unique mix of west African lowland and afromontane species (cf. Althof, 2005). Kakamega Forest once formed a single forest block together with the two Nandi Forests. However, since 1912/13 it has lost about 60% of its forest cover (Mitchell *et al.*, 2006) revealing a rather degraded state today (Schaab *et al.*, 2010). The people living in close vicinity to the forest depend on the forest for satisfying their daily needs (Gaesing, 2009). Efforts in harmonizing its management by the two authorities Kenya Wildlife Service (KWS) and Kenya Forest Service (KFS) have resulted in a participatory forest management plan including BIOTA outcomes (Mitchell *et al.*, 2009) and being ready to be launched in early 2011.
The farmland receives high annual rainfall suggesting a good agricultural potential and can be subdivided into the so-called sugarcane (in the north) and tea (in the south) zones. Subsistence farming is prevalent with a biannual harvest of a maize-bean intercrop (Jätzold *et al.*, 2005). For generating income, farmers also plant sugarcane and tea; other sources of income include informal employment and remittances (Gaesing, 2009). Because household earnings are generally so low, a diversification of income sources is required (Rietdorf, 2009). Commercial large-scale farming is hardly found in the area with the exception of a tea estate. Due to the ever-increasing demand for farmland by its growing population the area surrounding Kakamega Forest is already today highly structured (see Figure 1).

Figure 1. The highly structured farmland north of Kakamega Forest, western Kenya (by T. Lübker).

Very high resolution QuickBird satellite imagery (MS bands 2.40 m, PAN 0.60 m, acquired Feb/Mar 2005) covering 244 km² forest and 473 km² farmland was purchased to serve the interdisciplinary group of project partners with detailed information. While for the forest reserve a vascular plant community map has been derived via a visual interpretation based on vegetation surveys and in situ knowledge (Schaab *et al.*, 2010), the imagery of the highly structured farmland required an automated object-based approach. It has been anticipated to use the results on land use and landscape structures for analyzing biological and socio-economic field findings by the partners, thus allowing recommendations for a sustainable land use planning with the target of an improved rural livelihood while taking pressure from the forest and its biodiversity (Lübker, and Schaab, 2006).

2. OBJECT-BASED IMAGE ANALYSIS

2.1 Preprocessing and data preparation

Atmospheric and orographic correction was carried out with ATCOR 3 and improved image quality and the comparability of the two image swaths considerably. In particular orographic effects could be minimized successfully, while the removal of a thin haze layer could not be achieved. For mosaicing the two

* Corresponding author.

already georeferenced 'standard' image swaths a procedure was elaborated with a relative geometrical adjustment of the two swaths in only a very narrow strip to avoid resampling for most of the data (Lübker, and Schaab, 2008b). In a combined qualitative and quantitative assessment of various pan-sharpening methods, the high-pass filter method (Erdas Imagine) was evaluated as being best for the imagery at hand and for the anticipated tasks (Lübker, and Schaab, 2008c).
In Oct. 2007 a ground truthing was conducted in twelve test areas of 2 or 2.25 km² selected based on eleven criteria to represent the highly heterogeneous agricultural matrix and to consider areas of BIOTA activities. Beforehand 411 interesting, unclear or characteristic features to be visited were marked of which only 8% could not be verified. Supported by a field assistant, in total 636 samples were recorded, 224 regarding structural elements, 389 regarding land use, and 23 with other comments (Lübker, and Schaab, 2008a). Building upon the knowledge gained and the information obtained during field verification, a detailed visual interpretation for five of the twelve test areas was performed (for an example subset see Figure 2, left). On-screen digitizing at a scale of approx. 1:1,000 resulted in more than 16,000 polygons (Lübker, and Schaab, 2009), thus presenting a solid base for an evaluation of the segmentation quality later.

2.2 Segmentation

Adding the concept of groups (4) for a multilevel image segmentation, the classes as defined and adjusted for ground truthing (28) and visual interpretation (27) had to be rearranged, now considering in total 19 classes (Lübker, and Schaab, 2009). Since in object-based image analysis (OBIA) all subsequent classification steps depend on the quality of the segmentation result, the choice for an optimal parameter setting is fundamental. Here, for the region-based 'multiresolution' segmentation, five degrees of freedom have to be considered: layers to be selected, the weighting of layers, scale parameter, shape factor, and compactness (Lübker, and Schaab, 2009). Instead of testing a selection of 2,400 parameter combinations for five focus study sites and four groups of classes, a methodology has been developed to effectively facilitate the segmentation step. The approach is based on the determination of the first three listed degrees of freedom, keeping the others constant (120 parameter combinations, 600 candidate segmentations), by applying a newly introduced 'area fitness rate' as a discrepancy method, which calculates the degree of overlap between each reference polygon of the visual interpretation and the associated candidate segments. The two remaining degrees of freedom were then optimized while already adopting the just determined values via an 'objective function' as a goodness method, which balances inner-segment homogeneity and intra-segment heterogeneity (for a detailed description see Lübker, and Schaab, 2009). With in most cases very similar results obtained for the five focus sites selected to account for the heterogeneity of the area, a cross-check on the reliability of the results was performed. The determined parameter settings were judged representative and could thus be applied to the complete imagery (Lübker, and Schaab, 2009).

2.3 Classification

For selecting the features of relevance to distinguish one class from the others, the Seath tool was used. Class separability was determined using 980 selected objects from the twelve test areas as reference and taking into account 69 different object characteristics. The ten best distinguishing features were noted and used in the subsequent development of a knowledge-based rule set (considering in parts statistical measures), which consists of in total 831 individual processes. Here, the dependency on trial-and-error is a major drawback, which led to long development times (Lübker, and Schaab, 2010). To conclude, 15 LUC classes (plus shadow) of the anticipated 18 classes (not considering shadows here) could be distinguished (see Figure 2, right). This was judged satisfactory as well as the fact that only about two thirds of the houses could be extracted. It is believed that the shortcomings have to be ascribed to the data and not to the approach taken.
For processing the complete geodataset, a splitting into 300 tiles was necessary (Lübker, and Schaab, 2010). Obvious classification errors of e.g. features to be joined were manually edited when stitching the tiles together again. A suitable representation of the entire classification results (i.e. of the more than 700,000 polygons) is possible between the scales of 1:5,000 and 1:25,000 depending on the purpose, leading to map sizes of approx. 7.6 m by 5.4 m and 1.5 m by 1.1 m respectively.

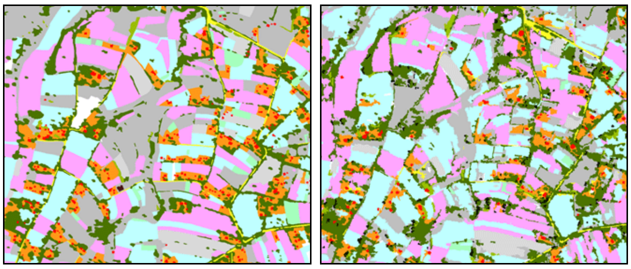

Figure 2. Visual interpretation (left) versus automated OBIA result (right) for a subset of the farmland in Buyangu village.

3. APPLICATIONS

3.1 Farmland typology

Due to the size and the level of detail, only an aggregation of the classification results enables to explore spatial patterns and thus to gain geospatial knowledge. Therefore, the farmland was divided into 1,324 hexagon-shaped areas (generally sized 41.67 ha). While straightforward maps of land use proportions already showed spatial patterns, three cluster analyses revealed more distinct spatial patterns. Based on the classification results, a DEM, as well as additional data retrieved through visual interpretation (i.e. schools and roads), for each individual hexagon meaningful landscape-determinant parameters of land use (6), landscape structures (4), and accessibility (3) were derived. Through hierarchical cluster analysis based on Ward's method three typology maps have been created, one for each topic. A synoptic cluster analysis based on all 13 parameters led to the final spatial farmland typology distinguishing ten different types of farmland (see Figure 3). The revealed ten types can serve as a planning basis for actions adjusted to the specific needs and problems of people living in similar settings (see Schaab et al., 2010).

3.2 Population redistribution

Census data assumes population density to be uniformly distributed within each administrative unit. To reflect the actual distribution, a GIS-based population surface model was developed (Ngochoch, 2007) redistributing human population in the Kakamega-Nandi forests area. Multiple geodatasets were employed for ancillary information, e.g. populated places, exclusion areas, infrastructures, rivers, and slopes. The ancillary

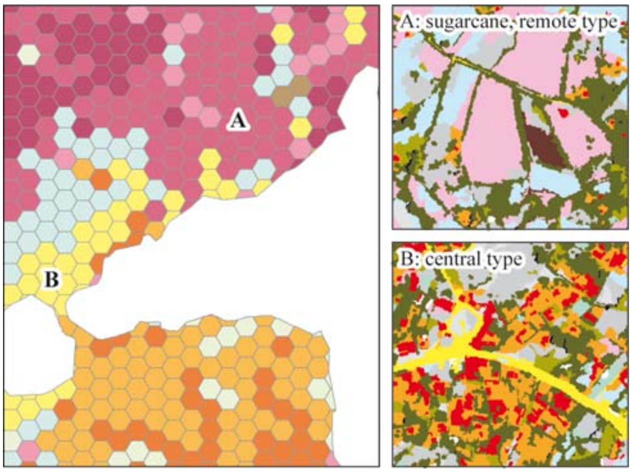

Figure 3. Subset of the Kakamega farmland typology with sample classifications to demonstrate the distinctness of types.

geodata was used for determining the most significant features affecting population distribution in each sub-region through logical correlation. The resulting 'population determinant factors' in turn impact on the weights given to the driving factors. Expertise showed to be very important in deciding on the logical correlations and weightings related to the ancillary data. Therefore, the original model was improved (Lung et al., in prep.) by using the houses as classified from the QuickBird satellite imagery. Polynomial frequency distribution functions of the houses related to slope, roads, rivers/streams, markets, and schools resulted in trustable weights per factor, which are applied now per 30 m pixel. Additionally, a mass preserving smoothing algorithm removes the abrupt changes in population density along the sublocation boundaries. The result (see Figure 4 or Schaab et al., 2010) is a more realistic pattern in the population distribution without obvious artefacts.

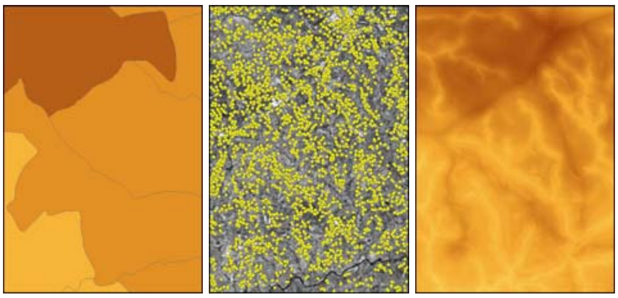

Figure 4. Comparing population density per administrative unit (left), houses as derived from QuickBird satellite imagery (centre), and model output of gridded population distribution (right).

3.3 Socio-economic scenarios

Both the typology and the QuickBird derived houses have been used to simulate alternative futures of rural livelihood, again with hexagons as the reference unit. In eight scenarios different developments in crop yields and prices were assumed for four time steps (2005-2020). Here, projections on population growth, an assumed decrease in household sizes, as well as additional area needed due to division of farms have been accounted for. The modelling output has been summarized for seven map topics, among them household production of important crops and household earnings through the sale of crops (see Figure 5). In total 175 thematic maps are available, which ask for a visualizing in a dynamic interactive tool to facilitate interpretation as well as local planning.

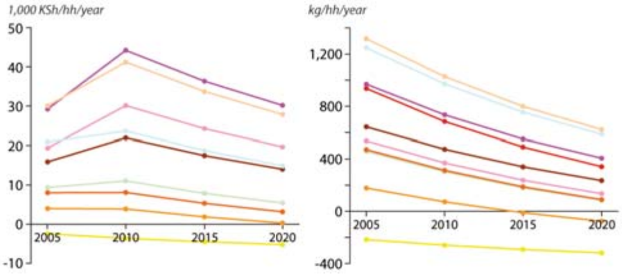

Figure 5. Diagrams visualizing per type of farmland possible future developments for two socio-economic topics.

3.4 Rainwater harvesting potential

In Kenya the area of Kakamega is known for plenty of rain. Recently and due to climate change impacts severe droughts hamper the availability of water and thus also the production of food. Therefore, it is striking that rainwater harvesting (RWH) from roof areas is not a common feature in the region. Concepts for a spatially explicit modelling of RWH potential in a GIS consider four different levels of detail and geographical extents, this dependent on available data of population distribution: census data and gridded population, houses as classified from the QuickBird data, and houses visually delineated from the same source covering just one village. Generally, the modelling procedure (Nthuni, 2010) covers the determination of spatial patterns of mean monthly rainfall, the rainwater endowment, the RWH potential, the water demand, and the rainwater use balance. Outputs aimed at are RWH potentials (Figure 6, left) or rainwater use balances, the determination of optimal tank sizes, and for the village level (see Figure 6, right) a comparison to collecting water from boreholes and streams.

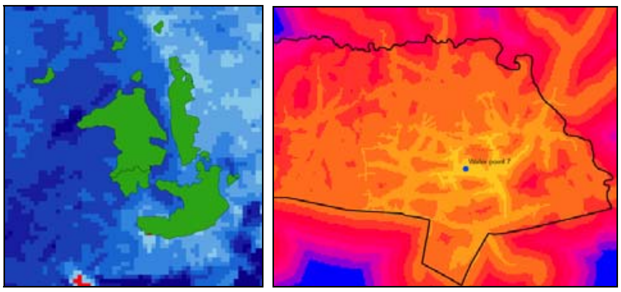

Figure 6. Modelling output on the annual RWH potential per person for the Kakamega-Nandi area (left), interim result of a cost surface related to water collection in Buyangu village (right).

4. CONCLUSIONS AND OUTLOOK

The processing of the QuickBird satellite image data turned out to have become a special challenge due to its large spatial extent (Lübker, and Schaab, 2009) with many difficulties studies often do not encounter due to employing small scene subsets only or aiming at the separation of a limited set of features. However,

although the satellite image analysis has taken several years, it has been worth the effort as demonstrated by the applications presented here. Changes in farmland use and structures are taking place continuously. Therefore, we consider an aggregation of the classification results a feasible and appropriate approach for planning purposes. This applies also to the subsequent simulation of socio-economic scenarios. For fine-tuning of model drivers or when considering only small areas, the classification results are an invaluable source for selecting the required features.
Such a wealth of detailed, spatially-explicit information has never before been available for the farmland studied. For its consideration in biodiversity-related land use planning efforts of BIOTA it may come too late. But many more applications for the benefit of the people and thus also of their forest are imaginable.

REFERENCES

Althof, A.J., "Human impact on flora and vegetation of Kakamega Forest, Kenya: Structure, distribution and disturbance of plant communities in an East African rainforest", PhD thesis at University of Koblenz-Landau (Germany), 2005.

Gaesing, K., "Reconciling biodiversity conservation with rural livelihood needs – A research concept under test in Kakamega Forest in Western Kenya", In: Gaesing, K. (Ed.), Reconciling Rural Livelihood and Biodiversity Conservation: Lessons from Research and Practice, SPRING research series, vol 52, p.p. 23-36, 2009.

Jätzold, R., H. Schmidt, B. Hornetz, and C. Shisanya, "Farm management handbook of Kenya. Vol II: Natural conditions and farm management information. Part A: West Kenya. Subpart A1: Western Province", 2nd ed., Ministry of Agriculture, and GTZ, Nairobi, 2005.

Kokwaro, J.O., "Conservation status of the Kakamega Forest in Kenya. The easternmost relic of the equatorial rain forests of Africa", Monographs in Systematic Botany (Missouri Botanical Garden), vol 25, p.p. 471-489, 1988.

Lübker, T., and G. Schaab, "Large area QuickBird imagery for object-based image analysis in Western Kenya: pre-processing difficulties, workarounds and resulting benefits as well as (first) segmentation strategies", ISPRS Proceedings of the 1st International Conference on Object-based Image Analysis (OBIA 2006) 'Bridging Remote Sensing and GIS', Vol. XXXVI – 4/C42 (digital), 4-5 July 2006, Salzburg (Austria), ed. by S. Lang, T. Blaschke, and E. Schöpfer, 2006.

Lübker, T., and G. Schaab, "'Ground truthing' – Erfassung von Referenzdaten für VHR-Satellitenbilddaten in einem ländlichen Gebiet Westkenias", In: Strobl, J., T. Blaschke, and G. Griesebner (Ed.), Angewandte Geoinformatik 2008, Beiträge zum 20. AGIT-Symposium in Salzburg, 2008, Heidelberg, p.p. 28-33, 2008a.

Lübker, T., and G. Schaab, "Identifying benefits of pre-processing large area QuickBird imagery for object-based image analysis", In: Blaschke, T., S. Lang, and G.J. Hay (Ed.), Object-based Image Analysis. Spatial Concepts for Knowledge-Driven Remote Sensing Applications, Berlin, Heidelberg, p.p. 203-219, 2008b.

Lübker, T., and G. Schaab, "Prozessierung von großflächigen QuickBird-Satellitenbilddaten für das Gebiet Kakamega Forest (Westkenia)", Photogrammetrie - Fernerkundung – Geoinformation (PFG), vol 1/2008, p.p. 15-27, 2008c

Lübker, T., and G. Schaab, "Optimization of parameter settings for multilevel image segmentation in GEOBIA", Proceedings of the ISPRS Hannover Workshop 2009 'High-Resolution Earth Imaging for Geospatial Information' (digital), 2-5 June 2009, Hannover; ed. by K. Jacobsen, S. Müller, and U. Sörgel, 2009.

Lübker, T., and G. Schaab, "A work-flow design for large-area multilevel GEOBIA: Integrating statistical measures and expert knowledge", Proceedings of GEOBIA 2010 – Geographic Object-Based Image Analysis, ISPRS Archives Vol.No. XXXVIII-4/C7, Vol. XXXVIII-4/C7 (digital), 29 June - 2 July 2010, Gent (Belgium), ed. by E.A. Addink, and F.M.B. Van Coilie, 2010.

Lung, T., J.K. Ngochoch, T. Lübker, and G. Schaab, "Calibrating regional population distribution models with VHR satellite imagery", Journal of Applied Earth Observation and Geoinformation, in prep.

Mitchell, N., T. Lung, and G. Schaab, "Tracing significant losses and limited gains in forest cover for the Kakamega-Nandi complex in western Kenya across 90 years by use of satellite imagery, aerial photography and maps", Proceedings of the ISPRS (TC7) Mid-Term Symposium 'Remote Sensing: From Pixels to Processes' (DVD), 8-11 May 2006, Enschede (The Netherlands), ed. by N. Kerle, and A.K. Skidmore, 2006.

Mitchell, N., G. Schaab, and J.W. Wägele (Eds), "Kakamega Forest ecosystem: An introduction to the natural history and the human context", Karlsruher Geowissenschaftliche Schriften (KGS), vol A 17, ed. by G. Schaab, 2009.

Ngochoch, J.K., "Gridding census population numbers for the BIOTA East Africa study areas – literature-based conceptualization and GIS-based implementation", Unpublished Masters thesis, study course Geomatics, Karlsruhe University of Applied Sciences (Germany), 2007.

Nthuni, S.M., "Spatial modelling of rainwater harvesting potential for the Kakamega area in western Kenya", Unpublished Masters thesis, study course Geomatics, Karlsruhe University of Applied Sciences (Germany), 2010.

Rietdorf, U., "Livelihood strategies of households in Kakamega District and their relation to biodiversity conservation", In: Gaesing, K. (Ed.), Reconciling Rural Livelihood and Biodiversity Conservation: Lessons from Research and Practice, SPRING research series, vol 52, p.p. 37-52, 2009.

Schaab, G., B. Khayota, G. Eilu, and J.W. Wägele (Ed.), "The BIOTA East Africa atlas. Rainforest change over time", Karlsruhe, 2010.

Schaab, G., T. Lübker, T. Lung, and N. Mitchell, "Remotely sensed data for sustainable biodiversity management. The case model of Kakamega Forest in western Kenya", Proceedings of the 33rd International Symposium on Remote Sensing of Environment 'Sustaining the Millennium Development Goals' (digital), 4-8 May 2009, Stresa, Lago Maggiore (Italy), ref. 479, 2009.

ACKNOWLEDGEMENTS

The work described has been achieved within the BIOTA East Africa project frame, funded by BMBF (FKZ 01LC0625D1).